



UNIVERSITY OF LEEDS

This is a repository copy of *Diagnosis With Nanoscale Protein Distributions: Single-Molecule Fluorescence Localization Microscopy and Attention-Based Learning*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/225595/>

Version: Accepted Version

Proceedings Paper:

Curd, A. orcid.org/0000-0002-3949-7523, Slaney, H., Brockmoeller, S. et al. (8 more authors) (Accepted: 2025) *Diagnosis With Nanoscale Protein Distributions: Single-Molecule Fluorescence Localization Microscopy and Attention-Based Learning*. In: 2025 IEEE International Symposium on Biomedical Imaging (ISBI). 2025 IEEE International Symposium on Biomedical Imaging (ISBI), 14-17 Apr 2025, Houston, TX, USA. IEEE (In Press)

This is an author produced version of an article accepted for publication in the proceedings of the 2025 IEEE International Symposium on Biomedical Imaging (ISBI), made available under the terms of the Creative Commons Attribution License (CC-BY), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

DIAGNOSIS WITH NANOSCALE PROTEIN DISTRIBUTIONS: SINGLE-MOLECULE FLUORESCENCE LOCALIZATION MICROSCOPY AND ATTENTION-BASED LEARNING

Alistair Curd^{1,2}, Hayley Slaney¹, Scarlet Brockmoeller^{1,3}, Andras-Gabor Miklosi⁴, Deep Arora⁵, Andrew Lewington⁶, James H. Felce⁴, Omar S.M. El Nahhas⁷, Marko van Treeck⁷, Jakob Kather⁷, Phil Quirke^{1,5}

¹ Leeds Institute of Medical Research at St James's, School of Medicine, University of Leeds, UK

² School of Molecular and Cellular Biology, Faculty of Biological Sciences, University of Leeds, UK

³ Departments of Histopathology, Queen Elizabeth Hospital, Gateshead and Precision and Molecular Pathology, Newcastle University, UK

⁴ ONI UK, Linacre House, Banbury Road, Oxford, UK

⁵ Department of Histopathology, Bexley Wing, St James's University Hospital, Leeds, UK

⁶ Renal Department, St. James's University Hospital, Leeds, UK

⁷ Else Kroener Fresenius Center for Digital Health, Technical University Dresden, Dresden, Germany

ABSTRACT

Single-molecule (fluorescence) localization microscopy (SMLM) finds the position of markers for target proteins at approx. 10 nm precision. Diagnosis of some diseases currently relies on inspection of nanoscale morphology by electron microscopy (EM), an expensive and slow test with limited sample coverage. Nanoscale biological processes also underlie health and disease in general, and so there is a need for more efficient diagnostic methods. We demonstrate that SMLM of routine biopsy samples can be used to assist diagnosis via data classification models. We predict diagnosis of 20 patients with chronic renal diseases (focal segmental glomerulosclerosis or minimal change disease) with a mean area under the receiver operating characteristic curve of 0.97 in cross-validation, and balanced accuracy of 90%. We tested state-of-the-art pretrained feature extraction from image tiles at 0.045 microns per pixel, followed by training of weakly supervised, attention-based models. SMLM and automated analysis has the potential to save time to diagnosis and costs compared with EM, with greater sample coverage, as well as for finding new nanoscale biomarkers in other disease areas.

Index Terms— *Single-molecule localization microscopy, renal disease diagnosis, weakly supervised learning, attention-based learning, nanoscale protein distribution*

1. INTRODUCTION

Information from immunohistochemistry, standard histopathology slides and confocal microscopy is limited by the diffraction of light to features larger than 200 nm in the best case. However, clinically important biological processes

take place at much shorter length scales. Therefore, electron microscopy (EM) is used when it is necessary to visualize nanoscale features to make a diagnosis. Medical EM is a specialized technique with specific sample preparation procedures and expensive equipment, in which only a small area of tissue can be selected for imaging.

A more recent form of fluorescence microscopy, single-molecule localization microscopy (SMLM) has the potential to assist diagnosis with nanoscale information on a tissue section [1]. SMLM fits into the standard sample preparation workflow for histopathology, and allows for scanning of a stained section of formalin-fixed paraffin-embedded (FFPE) tissue [2] and acquiring data from many regions of interest. SMLM finds the spatial positions of fluorescently labelled target molecules to a precision of approx. 10 nm and outputs a point cloud. Therefore, in principle it can be used as a faster, cheaper alternative to EM, with data from a larger area, and is specific to proteins of interest, while EM is non-specific.

SMLM and computer-assisted clinical decision making has not yet been demonstrated. The main use for SMLM has been in sub-cellular studies in biology, and imaging of routine FFPE tissue sections has been rare. Analysis methods for the data acquired in SMLM are also underdeveloped compared with other imaging techniques, despite much progress [3], including in machine learning methods [4, 5].

We demonstrate the use of SMLM in diagnosing two chronic renal diseases that currently require nanoscale imaging with EM: focal segmental glomerulosclerosis (FSGS) and minimal change disease (MCD). Incorrect diagnosis results in suboptimal treatment for both diseases, with a failure to appreciate the prognosis associated with these very distinct disease processes. In addition to a routine immunofluorescence panel (IF), EM is used in these diseases to show a loss of podocyte foot processes at the glomerular

basement membrane (GBM) and an absence of electron-density deposits. However, these techniques can be of limited value in FSGS cases when EM does not sample the focal nanoscale changes. We use SMLM data on sections from archived routine medical renal biopsies (FFPE tissue blocks), fluorescently stained for collagen IV, a GBM protein. Diagnosis by gold standard (IF and EM) is available as ground truth at the patient level for diagnosis predictions.

We predict FSGS or MCD diagnosis using image tiles constructed from the SMLM point cloud data, with multiple SMLM fields of view (FOVs, 50 x 80 μm each) per tissue section. Testing pretrained feature extractors [6, 7] and classification networks for tiled images (based on attention-based multiple-instance learning (attMIL) and vision transformer (ViT) [8-12]), we find good prediction of diagnosis. We also extract the most influential individual tiles in diagnosis prediction, for further use in investigating differences between the nanoscale protein distributions in the two diseases.

2. MATERIALS AND METHODS

2.1. Data acquisition

In sections from FFPE renal needle biopsy (one section per patient, 3 μm thick), we identified glomeruli and imaged 50 x 80 μm FOV glomerular regions with direct stochastic optical reconstruction microscopy (dSTORM), a type of SMLM [1, 13]. Collagen IV was stained with a primary antibody and a secondary antibody labelled with Alexa Fluor 647. As well as biopsies from renal patients diagnosed with FSGS or MCD, we imaged control sections from FFPE material from autopsy renal transplantation donors with no known renal condition, as a check on data acquisition quality (**Fig. 1**). Standard SMLM data processing steps resulted in the final point cloud of 2D collagen IV molecular positions, including drift correction over the five-minute acquisition time for a FOV and filtering for high-quality localizations [1]. We retained localizations with an estimated precision (s.d.) < 15 nm.

A 2D histogram of the localizations at 0.045 microns per pixel (mpp) gave an image to tile at 10 x 10 μm , 224 x 224 pixels (**Fig. 1D-F**). For effective training, we excluded tiles containing fewer than 1000 molecular localizations and subsequently excluded any FOVs from training and testing that had fewer than 8 tiles remaining. After filtering, the remaining data was from 400 FOVs across 20 patients out of 23 (3 sections had no FOVs remaining) (**Table 1**).

The number of FOVs available for an individual patient depended on the number of glomeruli captured by the 3 μm thick section of renal tissue and the area of the slice captured through a glomerulus (FOVs obtained from 2–28 glomeruli per section, median 10, after filtering as described; 1–4 FOVs available per glomerulus, mean 2.1). This resulted in a large range in the number of FOVs available for training and prediction per patient (2–41, median 19).

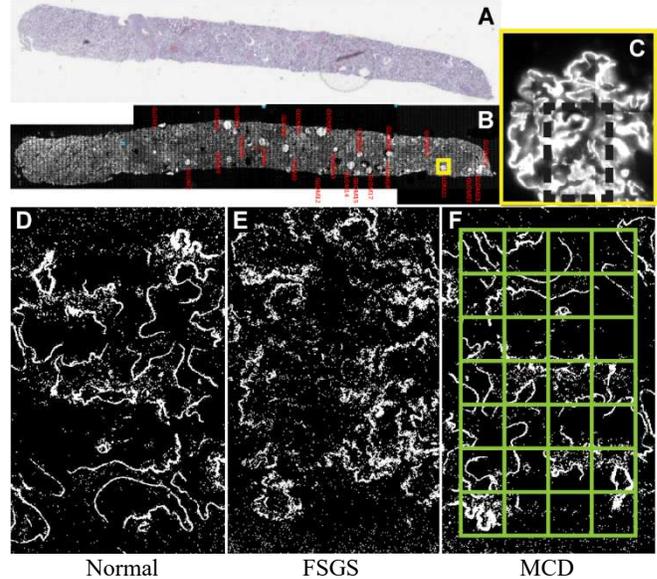


Fig. 1. Data acquisition and preprocessing. On a section of routine clinical material (**A**), fluorescence staining for glomerular proteins such as collagen IV (**B**) allows the location of glomeruli (box in **B**, **C**). SMLM data is acquired from 50 x 80 μm regions of a glomerulus (e.g. dashed black box in **C**), from sections from donors with no known condition (**Normal**, for control quality check) and from **FSGS** and **MCD** patients (**D-F**). 2D histograms of single collagen IV molecular positions (**D-F**) are tiled at 10 x 10 μm , 224 x 224 pixels (**F**). Images saturated here for visualization, but raw 2D histograms are used for feature extraction.

Table 1. Patient and SMLM FOV counts before and after filtering for higher-content tiles (> 1000 localizations) and FOVs (> 8 remaining tiles).

	FSGS		MCD	
	Patients	FOVs	Patients	FOVs
Unfiltered	11	222	12	351
Filtered	9	154	11	246

2.2. Feature extraction

We tested the embedding of tiles as feature vectors with pretrained networks, to subsequently feed into classification networks. DINOv2 is a self-supervised foundation model pretrained on 142M natural images [6]. UNI is an encoder that uses the DINOv2 model after fine-tuning on 100k whole-slide images from histopathology (WSIs) [7]. Our data is related to microscopic biological structures seen in histopathology, but may be out of domain for effective feature extraction with UNI, since resolution is $\sim 10\times$ higher than in the pretraining dataset (acquired at ~ 0.5 mpp [7]). There is also only a single information channel (collagen IV localization density), not RGB with a characteristic color spectrum as in WSIs. The fine-tuning of UNI may therefore be redundant, improve or worsen feature extraction versus

DINOv2, in our case. Thus, we tested both DINOv2-L/14 and UNI as feature extractors in our pipeline. We replicated the single-channel information in three channels, to mimic the dimensionality in their RGB training data.

2.3. Classification networks

We trained attMIL and ViT models previously optimized for weakly supervised learning, from tiled WSIs, of diagnosis and biomarkers for cancers [8, 10, 11]. The gold standard diagnosis for a renal biopsy (by IF and EM) labelled all tiles for that section, for model training and testing. We tested model performance for different batch sizes and selected the best parameters in cross-validation: batch size 16 for attMIL and batch size 1 for ViT, as previously found in its application to WSIs [11]. Training finished if there was no further improvement in validation loss over 16 consecutive epochs (patience), resulting in training over 77–142 (attMIL) and 8–89 (ViT) epochs in cross-validation experiments. With longer patience, the ViT tended to overfit to the training set.

2.4. Tasks and validation

We trained classification models on SMLM FOVs labelled with a diagnosis of FSGS or MCD, and tested their performance in two tasks: 1. Initial feasibility investigation of the pipelines in diagnosis prediction for a single FOV, without grouping the FOVs by patient; and 2. diagnosis prediction for a patient.

Task 1 used stratified 5-fold cross-validation of performance on the 400 FOVs, giving an estimate of the area under the receiver operating characteristic curve (AUROC) and balanced accuracy (BA) of predictions for the two diagnoses. We concatenated the test set prediction probabilities from all five folds as input to DeLong’s test on differences in AUROC between the pipelines.

In task 2, the set of FOVs from each patient was held out in turn as test data, in a 20-fold cross-validation. The training and validation sets were each allowed to contain FOVs from all other patients. The best model pipeline from task 1 was chosen to extract features and learn classification. The median confidence in FSGS and MCD diagnosis among the FOVs in the test set provided aggregated confidence values for the patient. We calculated BA of the 20 per-patient predictions, and estimated the model AUROC using 10,000 bootstrap samples of the 20 per-patient aggregated confidence values.

BA and recall calculations used a threshold of 0.5 for prediction of either diagnosis (FSGS or MCD).

2.5. Top tiles extraction

We ran tiles through the best model after training (highest AUROC), extracting the tiles ranked as the most influential in diagnosis prediction for the most confidently, correctly predicted patients. We also extracted the most influential tiles for the incorrectly predicted patients.

3. RESULTS

3.1. Testing pipelines on diagnosis prediction from single FOVs

Including all FOVs in 5-fold cross-validation without grouping by patient, mean AUROC for diagnosis prediction was above 0.88 for all feature extractors and trained classification models. Mean BA varied from 0.78 to 0.93 (Table 2). The best performing pipeline used tile feature extraction with UNI and prediction training with attMIL, resulting in mean AUROC 0.98 and BA 0.93.

Table 2. Performance of model pipelines in diagnosis prediction (FSGS or MCD) from single FOVs, not separated by patient. Mean and 95% confidence interval calculated from 5-fold cross-validation.

Pretrained feature extractor	Trained classifier	AUROC	Balanced accuracy (BA)
UNI	attMIL	0.979 ± 0.018	0.933 ± 0.024
DINOv2	attMIL	0.965 ± 0.007	0.893 ± 0.055
UNI	ViT	0.958 ± 0.030	0.845 ± 0.090
DINOv2	ViT	0.889 ± 0.049	0.778 ± 0.145

Training on tile features extracted with UNI, attMIL resulted in a greater AUROC and improved BA compared with ViT. The difference between attMIL and ViT results was statistically significant ($p < 0.001$, two-tailed DeLong’s test). Feature extraction with DINOv2 increased this performance difference between attMIL and ViT.

The UNI tile feature extractor resulted in higher estimates of AUROC and BA than DINOv2 trained on natural images. The difference between results with the two feature extractors was less statistically significant when used with attMIL ($p = 0.06$) than with ViT ($p < 0.001$).

3.2. Diagnosis prediction for patients

Training the attMIL model on tile features extracted with UNI, we obtained BA of 0.899 (18 out of 20 patients correct, Table 3). The bootstrapped AUROC estimate was 0.970 [0.879–1.000] (mean [lower–upper 95% confidence interval]).

Patients had between 2 and 41 FOVs available for diagnosis prediction. Excluding patients with fewer FOVs for prediction may improve recall and BA (Table 3). Therefore, conversely, taking further sections and obtaining further images may be useful in practice, to improve accuracy. Only one patient out of 13 with more than 10 SMLM FOVs after filtering for information content was misclassified, and with low confidence (median FSGS confidence: 0.56; diagnosis MCD).

We also consider excluding patients where a binomial test on the sample sizes (# FOVs in predicted class / total # FOVs for patient) indicates that we cannot be statistically confident that prediction by a majority vote among the FOVs

Table 3. Per-patient prediction performance, including a subset with the most SMLM FOVs available in the tissue section (> 10 FOVs available) and a subset with a statistically significant majority vote on prediction from the FOVs available ($p < 0.05$, binomial test).

Patients included	FSGS		MCD	
	<i>N</i>	Recall	<i>N</i>	Recall
All	9	89% (8 / 9)	11	91% (10 / 11)
> 10 FOVs available	5	100% (5 / 5)	8	88% (7 / 8)
Significant majority for prediction	3	100% (3 / 3)	9	100% (9 / 9)

reflects a true majority. A low number of total FOVs available or a balanced distribution of FOVs in the two prediction classes would cause a tissue section to fail this test (e.g. $p > 0.05$, one-tailed binomial test). Excluding patients at $p > 0.05$, we retain 12 patients, with 100% correct predictions. Use of such a criterion therefore improves confidence in model results for the retained samples and could be a helpful indicator of when further sections and data acquisition are needed in practice, or when results should be considered inconclusive.

3.3. Tile extraction

The most influential tiles in correct predictions appear to show thicker, more continuous collagen structures and patterns over broader areas for FSGS than for MCD (**Fig. 2**, correct predictions). Diffuse or punctate collagen patterns occurring in MCD may lead to an incorrect FSGS prediction in this pipeline (**Fig. 2**: Predicted FSGS; diagnosis MCD). Other patterns may be found to influence the model result in downstream analysis of highly predictive tiles, including the size, shape and spacing of small collagen patches.

4. DISCUSSION AND CONCLUSION

We have demonstrated that SMLM can be used to predict diagnosis from the nanoscale distribution of specific proteins in clinical FFPE tissue samples. We used tiled SMLM FOVs ($50 \times 80 \mu\text{m}$, 0.045 mpp) acquired from chronic renal disease biopsies (FSGS and MCD), which are currently assessed by EM for nanoscale morphological features of the glomeruli and podocytes. Over 20 patients, we obtained a mean AUROC estimate of 0.97 and balanced accuracy of 0.90 for diagnosis, using a pretrained feature extractor, trained on tiled WSIs [7], and attMIL [8, 10]. Future research will include the assessment of confidence in diagnostic prediction and practical decisions on when to acquire extra data for patients, which takes five minutes per FOV.

Both attMIL and ViT with pretrained feature extractors have previously been applied to classification and regression of tiled WSIs [11, 14, 15]. We found that the UNI feature

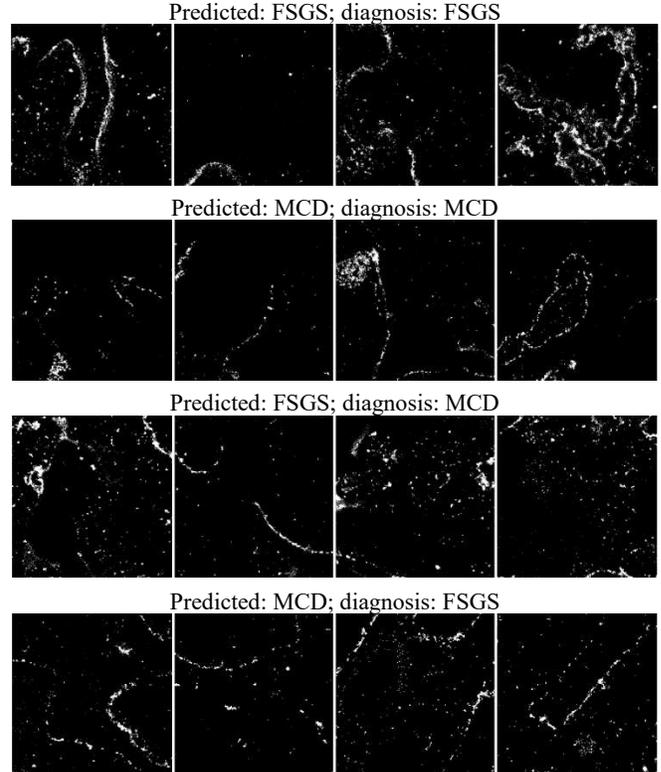


Fig. 2. Most confidently predicted $10 \times 10 \mu\text{m}$ tiles, from the most confidently predicted patients, for correct (top) and incorrect (bottom) predictions.

extractor, pretrained on WSIs at ~ 0.5 mpp [7], extracts relevant features for classification outside of its training domain, from our greyscale images of molecular density at a much higher resolution of 0.045 mpp. Feature extractors pretrained on fluorescence microscopy data are desirable and may emerge from the development of non-clinical models [16, 17].

Diagnosis prediction was more accurate with attMIL than with ViT, but only 400 FOVs from 20 patients were available for training and testing. We expect a larger cohort to improve both attMIL and ViT performance in a future validation study. Combining SMLM data on additional proteins (including nephrin in these renal diseases) and experimentation with resolution of the rendered FOVs may also improve model performance.

SMLM and prediction models have the potential to assist in diagnosis as a faster, cheaper, and protein-specific alternative to EM. Nanoscale protein distributions from SMLM also hold new data for understanding markers and underlying processes in disease, including in disease areas where EM is not currently considered. As such, our study points towards the value of SMLM and automated data analysis as a new technique in routine pathology.

5. COMPLIANCE WITH ETHICAL STANDARDS

This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted from the NHS Health research Authority, Leeds Teaching Hospital Trust, University of Leeds (IRAS Project ID: 270613 including two amendments).

6. ACKNOWLEDGMENTS

This study was funded by an NIHR i4i grant (NIHR201643) to ONI Ltd. and University of Leeds. AC was partially funded by Wellcome (204825/Z/16/Z). SB was funded through an NIHR Clinical Lecturer Fellowship and a Jean Shanks Clinical Lecturer fellowship grant. Phil Quirke is an NIHR Senior Investigator Emeritus and is supported in part by the NIHR Leeds Biomedical Research Centre (NIHR203331). Views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. JHF and AGM are employees of ONI Ltd. For the purpose of Open Access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

AC thanks Derek Magee for advice on computational experiments.

7. REFERENCES

- [1] M. Lelek *et al.*, "Single-molecule localization microscopy," *Nature Reviews Methods Primers*, vol. 1, article no. 39, June 2021.
- [2] L.C.S. Wunderlich, F. Ströhl, S. Ströhl, O. Vanderpoorten, L. Mascheroni, and C.F. Kaminski, "Superresolving the kidney—a practical comparison of fluorescence nanoscopy of the glomerular filtration barrier," *Analytical and Bioanalytical Chemistry*, vol. 413, no. 4, pp. 1203–1214, February 2021.
- [3] S. Liu, P. Hoess, and J. Ries, "Super-Resolution Microscopy for Structural Cell Biology," *Annual Review of Biophysics*, vol. 51, pp. 301–326, May 2022.
- [4] I.M. Khater, I.R. Nabi, and G. Hamarneh, "A Review of Super-Resolution Single-Molecule Localization Microscopy Cluster Analysis and Quantification Methods," *Patterns*, vol. 1, no. 3, article no. 100038, June 2020.
- [5] S. Hugelier *et al.*, "ECLiPSE: a versatile classification technique for structural and morphological analysis of 2D and 3D single-molecule localization microscopy data," *Nature Methods*, vol. 21, no. 10, pp. 1909–1915, October 2024.
- [6] M. Oquab *et al.*, "DINOv2: Learning Robust Visual Features without Supervision," *arXiv:2304.07193*, Feb 2024.
- [7] R.J. Chen *et al.*, "Towards a general-purpose foundation model for computational pathology," *Nature Medicine*, vol. 30, no. 3, pp. 850–862, March 2024.
- [8] M. Ilse, J. Tomczak, and M. Welling, "Attention-based Deep Multiple Instance Learning," in *Proceedings of the 35th International Conference on Machine Learning*, PMLR, Stockholm, 2018, vol. 80, pp. 2127–2136.
- [9] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in *9th International Conference on Learning Representations*, OpenReview.net, Vienna (Virtual), 2021.
- [10] O.L. Saldanha *et al.*, "Self-supervised attention-based deep learning for pan-cancer mutation prediction from histopathology," *npj Precision Oncology*, vol. 7, article no. 35, March 2023.
- [11] S.J. Wagner *et al.*, "Transformer-based biomarker prediction from colorectal cancer histology: A large-scale multicentric study," *Cancer Cell*, vol. 41, no. 9, pp. 1650–1661, September 2023.
- [12] O.S.M. El Nahhas *et al.*, "From whole-slide image to biomarker prediction: end-to-end weakly supervised deep learning in computational pathology," *Nature Protocols*, vol. 20, pp. 293–316, January 2025.
- [13] M. Heilemann *et al.*, "Subdiffraction-resolution fluorescence imaging with conventional fluorescent probes," *Angewandte Chemie*, vol. 47, no. 33, pp. 6172–6176, July 2008.
- [14] M.Y. Lu, D.F.K. Williamson, T.Y. Chen, R.J. Chen, M. Barbieri, and F. Mahmood, "Data-efficient and weakly supervised computational pathology on whole-slide images," *Nature Biomedical Engineering*, vol. 5, no. 6, pp. 555–570, June 2021.
- [15] O.S.M. El Nahhas *et al.*, "Regression-based Deep-Learning predicts molecular biomarkers from pathology slides," *Nature Communications*, vol. 15, article no. 1253, February 2024.
- [16] H. Kobayashi, K.C. Cheveralls, M.D. Leonetti, and L.A. Royer, "Self-supervised deep learning encodes high-resolution features of protein subcellular localization," *Nature Methods*, vol. 19, no. 8, pp. 995–1003, August 2022.
- [17] C. Ma, W. Tan, R. He, and B. Yan, "Pretraining a foundation model for generalizable fluorescence microscopy-based image restoration," *Nature Methods*, vol. 21, no. 8, pp. 1558–1567, August 2024.